

Miguel, Sandra; Caprile, Lorena; Jorquera Vidal, Israel

Análisis de co-términos y de redes sociales para la generación de mapas temáticos de la colección

V Jornadas de Sociología de la UNLP

10, 11 y 12 de diciembre de 2008

Cita sugerida:

*Miguel, S.; Caprile, L.; Jorquera Vidal, I. (2008). Análisis de co-términos y de redes sociales para la generación de mapas temáticos de la colección. V Jornadas de Sociología de la UNLP, 10, 11 y 12 de diciembre de 2008, La Plata, Argentina. En Memoria Académica. Disponible en:
http://www.memoria.fahce.unlp.edu.ar/trab_eventos/ev.6251/ev.6251.pdf*

Documento disponible para su consulta y descarga en **Memoria Académica**, repositorio institucional de la **Facultad de Humanidades y Ciencias de la Educación (FaHCE)** de la **Universidad Nacional de La Plata**. Gestionado por **Bibhuma**, biblioteca de la FaHCE.

Para más información consulte los sitios:

<http://www.memoria.fahce.unlp.edu.ar> <http://www.bibhuma.fahce.unlp.edu.ar>



Esta obra está bajo licencia 2.5 de Creative Commons Argentina.
Atribución-No comercial-Sin obras derivadas 2.5

Análisis de co-términos y de redes sociales para la generación de mapas temáticos de la colección

Sandra Miguel¹, Lorena Caprile² e Israel Jorquera Vidal²

Resumen

La mejora de las capacidades de búsqueda y de las interfaces de los OPACs sigue siendo uno de los principales desafíos para las bibliotecas, especialmente en lo que respecta al acceso por materias. Las interfaces visuales pueden ser una alternativa para facilitar la recuperación. El objetivo del presente trabajo es explorar si la combinación de técnicas de análisis de co-términos y de redes sociales resulta ser una metodología válida para la generación de mapas temáticos de la colección. La principal conclusión es que el método es válido, y que los mapas obtenidos podrían servir tanto como una alternativa de interfaz visual para el acceso por materias, como también para detectar problemas en los criterios de indización, y contribuir a la mejora de la calidad de la descripción temática del conjunto documental.

Palabras claves

Análisis de redes sociales; Análisis de co-términos; OPACs; Visualización de información; Interfaces visuales; Recuperación de información

1. Introducción

Las investigaciones sobre OPACs durante los últimos veinte años revelan una incesante búsqueda por encontrar las mejores formas de representar la información, en tanto interfaces entre los usuarios y la literatura (**White y McCain, 1997**). Los principales desafíos están vinculados tanto con la mejora de las capacidades de búsqueda de los sistemas de recuperación de información, como con el diseño de sus interfaces (**Mi, 2008**).

A los nuevos desafíos se suman antiguos problemas aún no resueltos, como los relacionados con la búsqueda por materias (**Drabenstott y Weller, 1996**). Conforme a los resultados de algunos estudios, este tipo de búsqueda es la más solicitada por los usuarios, y paradójicamente es la que mayores dificultades plantea (**Larson, 1991; Ríos García, 1991; Su, 1994; Pichinini y Martínez, 1996**). En el 40% de los casos el fracaso en las búsquedas se relaciona con la sobrecarga de resultados, o por el contrario, con resultados nulos, por la falta de coincidencia entre los términos empleados por el usuario en el momento de la

1 Departamento de Bibliotecología, Facultad de Humanidades y Ciencias de la Educación, Universidad Nacional de La Plata, Argentina; Grupo SCImago, Universidad de Granada, - UA CSIC, España. Correo-e: sandra@fcnym.unlp.edu.ar

2 Biblioteca Facultad de Ciencias Naturales y Museo, Universidad Nacional de La Plata, Argentina. Correo-e: lorenacaprile@yahoo.com.ar; jorqueravidal@gmail.com.ar

recuperación, y los almacenados en el sistema. Ante este problema, un importante porcentaje de usuarios ha expresado su deseo de ver palabras relacionadas con su expresión de búsqueda.

Ello nos lleva a pensar que una interfaz que ofrezca al usuario la posibilidad de visualizar las materias del OPAC, podría ser una buena alternativa para reducir el fracaso en las búsquedas, especialmente cuando éste se debe a la falta de equivalencia entre los términos.

Cabe recordar que la interfaz usuario-sistema es el lenguaje de comunicación entre el hombre y la máquina (**Matthews**, 1983). Su propósito es facilitar la tarea de recuperación de información, reduciendo los errores humanos y mejorando el desempeño del usuario durante el proceso de la búsqueda de información.

De acuerdo a la clasificación propuesta por **Hildrech** (1985), la segunda generación de OPACs incorporó muchas funcionalidades y elementos de diseño de la interfaz orientados a mejorar las prestaciones de búsqueda por materia. Algunos ejemplos son la posibilidad de realizar consultas por campos y la inclusión de índices de materias. Se ha demostrado, que cuando estos índices tienen un mantenimiento regular y el usuario conoce bien su manejo, constituyen una herramienta de gran utilidad para obtener resultados exitosos (**Villén**, 2006).

De igual modo, y a pesar de estos avances, algunos autores sostienen que las interfaces de los OPACs aún deben dar un salto cualitativo y cuantitativo (**Játiva**, 2004) para conseguir mayor usabilidad.

Es sabido que la manera que se elija para presentar al usuario final la información es tan importante como el contenido, ya que la forma en la que se haga accesible en pantalla va a determinar que la información cumpla su finalidad de ser útil, o por el contrario se quede almacenada sin más (**Marcos**, 2003).

Particularmente, cuando el usuario no tiene claro lo qué está buscando, o cuando tiene dificultades para formalizar su necesidad de información a través del lenguaje de interrogación del sistema, se requiere de un modelo alternativo o complementario que le brinde otra vía de acceso al conocimiento. Este modelo alternativo, conocido como exploración o navegación (*browsing*), da al usuario la posibilidad de explorar o inspeccionar

el conjunto documental sin tener que expresar de forma previa y con una o más palabras cuales son esas necesidades (**Herrero-Solana y Hassan-Montero**, 2006).

En la mayoría de los OPACs actuales, las materias que representan temáticamente la colección se presentan al usuario como un listado de términos dispuestos en orden alfabético. En el mejor de los casos, se han incorporado ficheros de autoridades o tesauros. En estos últimos, la interfaz adopta, por lo general, la forma de un directorio jerárquico.

Ejemplos son *MeSHBrowser* (<http://www.nlm.nih.gov/mesh/MBrowser.html>), que permite la navegación jerárquica entre los encabezamientos del *Medical Subject Headings* (**Korn**, 1997), y *DeweyBrowser*, una interfaz que utiliza la clasificación temática de Dewey para facilitar al usuario la navegación entre las clases y subclases del esquema (<http://deweybrowser.oclc.org/>).

La superación de estas estrategias de exploración viene de la mano de las investigaciones sobre Interfaces Visuales de Recuperación de Información (VIRI). El principal objetivo de estas interfaces es la representación visual e interactiva de la información en un espacio de dimensiones reducidas, resultado de la explotación de los datos (*data mining*) y de la abstracción gráfica del conjunto documental.

Para **Noyons y otros** (1999) el uso de mapas como metáforas de un campo de conocimiento tiene un gran potencial como interfaz para la recuperación de la información. Por su parte, **Janecek** (2005) plantea que las interfaces visuales son potencialmente útiles para que los usuarios puedan explorar una representación de la colección, y oportunamente descubrir información que los guíe a encontrar los documentos para ellos relevantes.

Desde esta perspectiva, la interacción con representaciones visuales de los esquemas de clasificación temática, debería permitir a los usuarios tener un mapa mental de cómo está organizada la biblioteca, lo que suponemos haría más fácil el acceso por materias.

Cabe señalar, que la visualización de información no es el resultado implícito del acto de ver. Requiere transformar datos abstractos y fenómenos complejos de la realidad en mensajes visibles, haciendo posible que los individuos vean con sus propios ojos datos y fenómenos que yacen ocultos y que no son directamente aprehensibles. En otras palabras, la

visualización consiste en hacer visible para nuestra mente aquello que no es visible para nuestros ojos (**Vargas Quesada y Moya Anegón**, 2007).

2. Iniciativas para el desarrollo de interfaces visuales de los sistemas de recuperación de información

En los últimos años varios autores se han dedicado a estudiar las estrategias e iniciativas de desarrollo de interfaces visuales. **Moya Anegón y Herrero-Solana** (1999) analizaron varios proyectos de investigación sobre interfaces gráficos, bidimensionales y tridimensionales, que permitieran la exploración de grandes volúmenes de información. **Marcos** (2004) realizó una revisión y descripción detallada de varias iniciativas, diferenciándolas según estuvieran orientadas a la interfaz de acceso o a la forma de presentación de los resultados. **Hassan-Montero y Herrero-Solana** (2006) analizaron diferentes estrategias para el desarrollo de interfaces visuales, todas ellas basadas en la identificación de las relaciones semánticas de similitud existentes entre documentos, o incluso entre términos. Asimismo, dos excelentes revisiones sobre técnicas de visualización son las de **White y McCain** (1997) y **Börner y otros** (2003). Nosotros hacemos mención de algunas de las iniciativas, en un intento por mostrar la diversidad metodológica de las diferentes propuestas.

Entre estas propuestas están los mapas auto-organizativos (*self-organizing map* o *SOM*). Estos mapas están basados en un tipo de red neuronal que trabaja bajo el principio de auto-organización y clasificación automática de la información. Cada nodo en el mapa está asociado con una lista de documentos semánticamente similares, y la proximidad espacial entre ellos indica una similitud conceptual (**Börner**, 2003). Las primeras aplicaciones al campo de la recuperación de información se atribuyen a Xia Lin y otros (**Lin, Soergel, & Marchionini**, 1991), y dos de los desarrollos prototípicos más conocidos son WEBSOM, dirigido por Kohonen (websom.hut.fi/) y ET-Map dirigido por Chen. Sin embargo, no hemos encontrado un sistema que actualmente utilice estos mapas como interfaz de recuperación. Algunos autores han señalado que su construcción exige una gran capacidad de procesamiento, y que presenta limitaciones en cuanto a la cantidad de información que puede procesar (**Guerrero Bote**, 2002).

Otros proyectos apuestan a la generación de interfaces utilizando técnicas de análisis de cluster (*clustering*). Con estas técnicas es posible hacer agrupamientos de documentos en función de patrones de proximidad o distancia, obteniendo como resultado clases o *clusters* de documentos similares. Una de las más recientes iniciativas basada en estos principios es Grokker (www.grokker.com/), un buscador web cuya interfaz visual permite al usuario navegar entre los *clusters*, e ir seleccionando dentro de ellos los enlaces de su interés. A pesar de que es visualmente atractivo, tiene como desventaja que a medida que se va descendiendo en niveles de agregación más específicos, se va perdiendo la vista del todo, aspecto que a nuestro juicio produce cierta desorientación.

Autores como **Chen** (1998) y **White** (2003), entre otros, propusieron el uso de redes *pathfinder* (PFNET) para la navegación visual. La premisa básica de la que parten estos autores, es que los usuarios de los sistemas de recuperación de información, necesitan conocer como está organizado un ambiente para poder utilizar el conocimiento y encontrar lo que buscan. Desde esta perspectiva, el diseño de interfaces requiere asociar las relaciones semánticas implícitas entre los documentos que conforman la colección de un dominio específico, y las necesidades cognitivas de los usuarios.

Otras propuestas de modelos de interfaces surgen de la combinación de varios métodos: minería de metadatos, escalamiento multidimensional y una técnica de *focus+context* denominada “ojo de pez” (**Hassan Montero y Herrero Solana**, 2006). Lo interesante de esta propuesta radica en que los autores desarrollaron un prototipo y realizaron una prueba de uso con usuarios reales, quienes valoraron de forma positiva la interfaz visual propuesta.

Otro tipo de representación visual que ha cobrado especial énfasis en los últimos años es la basada en la teoría de grafos. Un grafo es un conjunto de nodos con unas relaciones explícitas entre ellos (**Caraballo Pérez**, 2007). Los nodos representan actores (documentos, autores, palabras, términos, etc.), y los enlaces relaciones de algún tipo. Una representación de este tipo es la que presenta un *Topic Maps* (<http://www.topicmaps.org/>).

Un *Topic Maps* tiene como objetivo estructurar la información y presentarla en la forma de una red semántica. Un ejemplo de interfaz con estas características es la que usa *Aquabrowser* (www.aquabrowser.com/), un sistema que permite visualizar, rápida y

organizadamente, un conjunto de términos relacionados conceptualmente con los propuestos por el usuario (**Morillo-Velarde Serrano**, 2006; **Senso**, 2006). Lo interesante de este sistema es que no es un prototipo, como la mayoría de los desarrollos existentes hasta ahora, sino que está en funcionamiento en muchas bibliotecas. Dos ejemplos lo constituyen Queens Library (<http://aqua.queenslibrary.org/>) y Arlington Public Library (<http://www.acornweb.org/>).

De igual modo, *Aquabrowser* no permite obtener una representación temática sintética de la colección de la biblioteca. Hace ya más de diez años, **Schneiderman** (1996) señalaba que las interfaces de los sistemas de recuperación de información deberían proveer una primera representación gráfica o mapa, que sea lo suficientemente general e informativo para que el usuario pueda tener una idea global del dominio de conocimiento en cuestión, para luego ir haciendo *zoom* y descender a niveles cada vez más específicos.

Un ejemplo de sistema que utiliza una interfaz visual que aplica esta estrategia para la representación de dominios científicos es el Atlas de la Ciencia (www.atlasofscience.net/), desarrollado por el Grupo SCImago. La denominación de Atlas alude a un conjunto de mapas que son utilizados como interfaz, a través de la cual el usuario puede navegar entre categorías temáticas, en distintos niveles de agregación, hasta llegar a los documentos propiamente dichos (**Moya Anegón y otros**, 2004). En este caso, la combinación de técnicas de análisis de co-citación y de redes sociales ha demostrado ser una estrategia válida para la generación automática de estas interfaces visuales e interactivas.

El objetivo de este trabajo es explorar si la combinación de técnicas de análisis de co-términos y de redes sociales es válida para la generación de mapas temáticos de las colecciones de las bibliotecas, y presentar sus principales ventajas y limitaciones.

3- Análisis de co-términos

Por co-términos entendemos la co-ocurrencia o aparición conjunta de dos términos en un corpus textual dado. El análisis de co-términos es una variante del análisis de co-palabras. La diferencia radica en que los términos hacen referencia a representaciones de conceptos especializados en uno o más dominios de conocimiento, y las palabras aluden al vocabulario general (**Arntz**, 1995).

En cualquier caso, se trata de una técnica de análisis de contenido que usa patrones de co-ocurrencia de pares de ítems (palabras o términos) dentro de un corpus textual, para identificar las relaciones entre las ideas en un dominio determinado. El corpus, en nuestro caso, es cada registro bibliográfico del OPAC, y los términos las expresiones utilizadas para la descripción del contenido de los documentos, y que son incluidos en los campos de materia de la base de datos. Desde esta perspectiva de análisis, dos términos (y por extensión dos documentos) estarán más vinculados o asociados conceptualmente entre sí, cuanto mayor sea la co-ocurrencia entre ellos.

La técnica de análisis de co-palabras fue desarrollada en los años ochenta por Michelle Callon y Jean Pierre Latour (**Callon**, 1986; **Latour**, 1992). Aunque fue primeramente propuesta con la finalidad de mapear la dinámica de los campos científicos, el propio Callon señaló su utilidad como un método para clasificar documentos con fines de recuperación de información, y construcción y actualización de tesauros (**He**, 1999). Más recientemente, algunos autores propusieron su uso para la creación de interfaces de bibliotecas digitales (**Buzydlowski y otros** 2002); (**Börner y Chen**, 2002).

4- Análisis de redes sociales

El análisis de redes sociales es una nueva perspectiva conceptual y metodológica del campo de las ciencias sociales y del comportamiento, que a diferencia del enfoque tradicional de análisis atributivo de los elementos de un grupo, centra su atención en el análisis de las relaciones existentes entre esos elementos. Los datos son relacionales, entendiendo por dato relacional un vínculo específico existente entre un par de elementos. Desde esta perspectiva, la relación entre pares es una propiedad del par (es decir de la relación) y no una cuestión inherente a las características individuales de cada uno de los elementos relacionados (**Wasserman y Faust**, 1998).

Una red social está conformada por al menos dos componentes: los actores, que son las unidades que se relacionan entre sí por algún motivo, finalidad o circunstancia, y las relaciones, que son los lazos o vínculos existentes entre los actores.

Los actores de una red pueden desempeñar diferentes roles según sea su grado, poder de intermediación en relación a los demás, cercanía, etc. Asimismo, los vínculos entre ellos

pueden ser directos o indirectos, direccionales o no-direccionales, tener diferentes intensidades; aspectos todos que van determinando el tipo de relación existente y el tipo de estructura de red que conforman (**Rodríguez**, 1995). A cada actor de la red se lo identifica con un símbolo (generalmente un círculo), y su posición en la estructura está determinada por el rol que desempeña.

Las representaciones gráficas de las redes sociales son grafos. Como son relativamente fáciles de comprender, son ideales como metáfora visual para la representación de información (**Dürsteler**, 2003), además de que permiten descubrir relaciones que de otra forma permanecerían ocultas (**Vargas Quesada**, 2005).

Si partimos de la asunción de que la creación de conocimiento es una construcción social, donde un actor puede ser alguien o algo creador de asociaciones: un científico, un texto o un concepto, y donde las asociaciones o relaciones no son estáticas sino que cambian y se transforman constantemente (**Buchelli Guerrero**, 2008), una red social de términos de un campo temático podría representar la estructura de conocimiento de dicho campo.

Si trasladamos estas ideas a las bibliotecas, los documentos que integran la colección (y los términos que los describen temáticamente) conformarían un entramado o red, cuya evolución estaría determinada por la dinámica de su desarrollo; y cuya representación gráfica sería un buen reflejo de su organización temática.

5- Aplicación de las técnicas de análisis de co-términos y de redes sociales para la generación de mapas temáticos de la colección

Para determinar si la combinación de técnicas de análisis de co-términos y de redes sociales resulta válida para la generación de mapas temáticos de las colecciones, realizamos una prueba utilizando como fuentes de datos los OPACs de libros y revistas de dos bibliotecas universitarias de nivel de Facultad, dependientes de la Universidad Nacional de La Plata, Argentina. La razón por la que elegimos estas instituciones radica en que sus colecciones representan temáticas de campos disciplinares con patrones de producción de conocimientos diferenciados. Una se especializa en el área de las ciencias naturales (CNAT) y la otra en ciencias humanas y sociales (HUMS). Ello nos permite evaluar si la metodología es aplicable a colecciones de temáticas disímiles.

En total hemos analizado las colecciones representadas en cuatro catálogos: LCNAT (OPAC de libros de ciencias naturales); LHUMS (OPAC de libros de humanidades y ciencias sociales); RCNAT (OPAC de revistas de ciencias naturales) y RHUMS (OPAC de revistas de humanidades y ciencias sociales).

La población estudiada correspondió a la totalidad de los registros de los catálogos de revistas (5.249 para RCNAT y 3.002 para RHUMS), y al subconjunto de registros de los catálogos de libros que tenían completos los campos de materia (48% para LCNAT -9.350 de un total de 19.623 registros- y 45% para LHUMS, -29.115 de un total de 64.822-). De cada uno de los registros extraímos los términos que representan los temas de los documentos en dos niveles de agregación, uno más general y otro más específico. El primero se identifica con el nivel de disciplina (ej: Geología, Biología, Antropología, Historia, Literatura, etc.), y el segundo, con palabras claves de naturaleza y grado de especificidad diversa. Cabe señalar que HUMS utiliza un tesoro para la indización, y CNAT una lista de términos elaborada ad-hoc. En ambos casos hubo que realizar un proceso de normalización de las entradas.

Calculamos las frecuencias de ocurrencia de cada término y seleccionamos el subconjunto de aquellos con mayor presencia, debido a que una red con una cantidad excesiva de nodos y relaciones presenta serias dificultades de interpretación. La determinación de donde establecer el punto de corte no es tarea sencilla. Sin embargo, hay varios elementos a considerar. Uno, es que la cantidad de nodos no sea excesivamente grande para facilitar la clara identificación de cada uno de ellos. Otro, es no excluir nodos que actúan como puente de otros nodos. Estos reciben el nombre de puntos de corte o intermediarios, justamente porque quitarlos produce la desconexión de otros nodos de la estructura, y un incremento del número de componentes de la red. Considerando estos aspectos, hemos fijado en cada caso un umbral donde no solo se considere el peso individual de cada término, sino su relación con el resto de los términos que conforman la estructura.

Una vez seleccionados los términos construimos una matriz cuadrática y simétrica con las frecuencias de co-ocurrencia de cada uno con el resto, y generamos el mapa temático de cada colección analizada.

Para los conteos de co-ocurrencia hemos aplicado como criterio, que por cada término de nivel más general (equivalente al nivel de disciplina) aparezca siempre asociado uno de nivel más específico (nivel de palabra clave), siempre que estuvieran presentes en el mismo registro bibliográfico. Por ejemplo, si un registro del OPAC tiene asignado como término genérico Antropología, y como palabras claves Etnología y Folklore, las asociaciones que se establecen son: Antropología <> Etnología; Antropología <> Folklore y, Folklore <> Etnología. Lo que pretendemos con ello es que la red no solo refleje las relaciones existentes entre los términos más específicos, sino que también permita visualizar las jerarquías que se establecen entre los términos y que responden al típico esquema de organización del conocimiento de las colecciones de las bibliotecas, basado en sistemas jerárquicos como el de la Clasificación Decimal Universal (CDU), Clasificación de Dewey (CDD) u otros semejantes, en los que el conocimiento es dividido en clases y subclases siguiendo un modelo que va desde lo general a lo particular.

La construcción de las redes y las representaciones gráficas las realizamos utilizando el programa Pajek. Para una mejor visualización de la estructura de red empleamos el algoritmo Kamada Kawai (**Kamada-Kawai**, 1989), y a efectos de facilitar la interpretación y hacer más legible las relaciones entre los nodos (términos) eliminamos los enlaces (líneas) más débiles. El tamaño de los nodos representa el peso que cada término tiene en la red, y el color y grosor de las líneas un indicador de la intensidad de la relación.

6- Análisis e interpretación de los resultados

Como resultados del análisis obtuvimos cuatro redes que representan los mapas temáticos de las colecciones de las respectivas bibliotecas (Figs. 1, 2, 3 y 4). Los mapas presentan rasgos comunes como diferenciados. Entre los rasgos comunes podemos mencionar:

Fig. 1 Mapa temático de la colección de libros de ciencias naturales – LCNAT

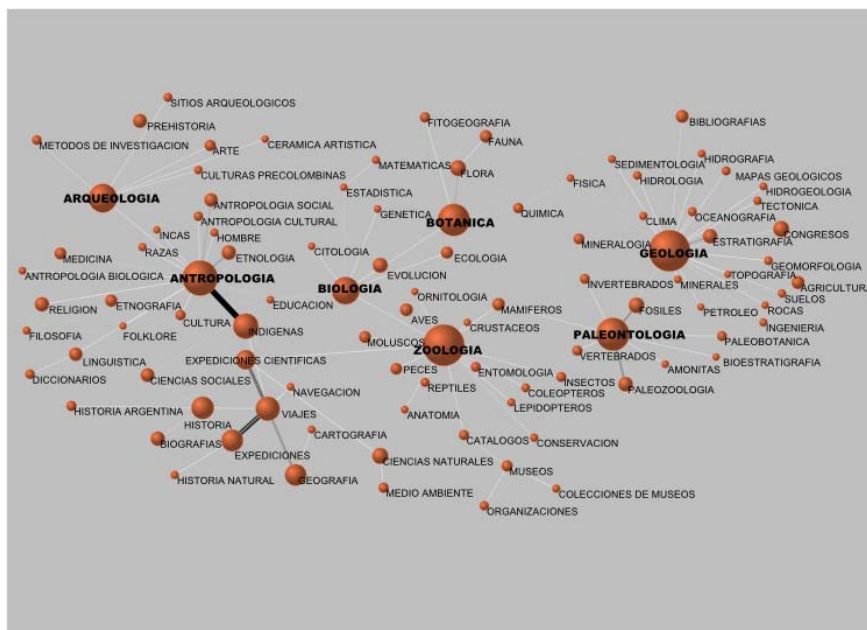


Fig. 2 Mapa temático de la colección de revistas de ciencias naturales – RCNAT

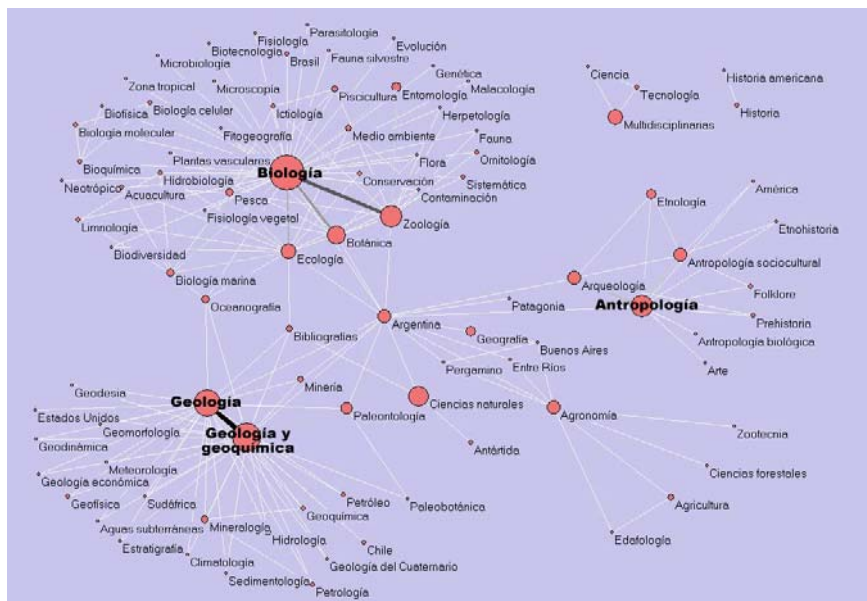


Fig. 3 Mapa temático de la colección de libros de humanidades y ciencias sociales – LHUMS

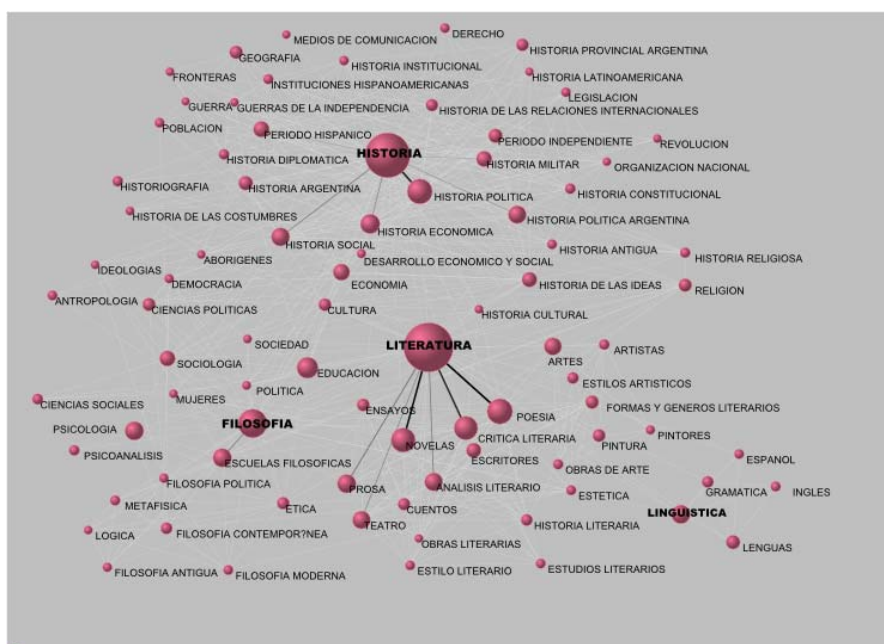
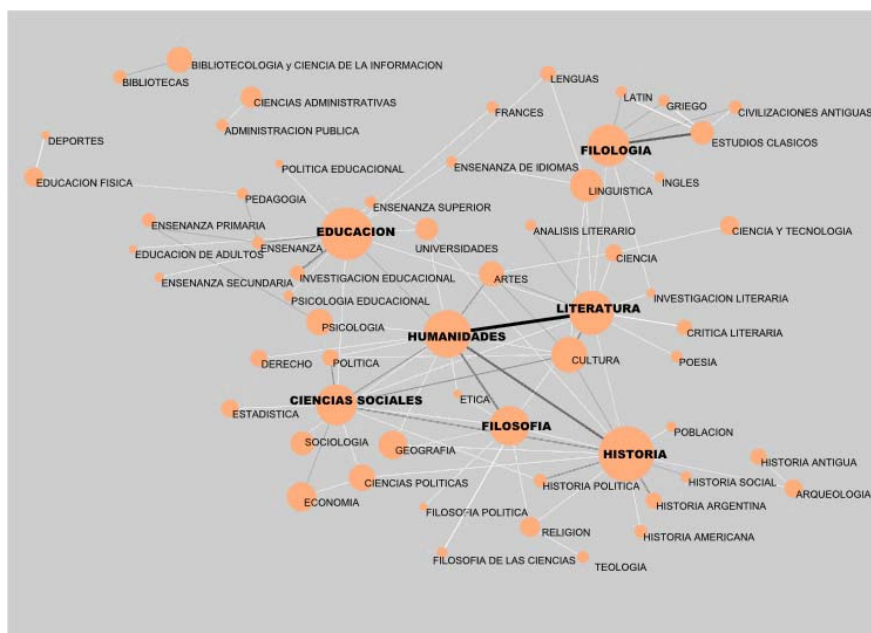


Fig. 4 Mapa temático de la colección de revistas de humanidades y ciencias sociales – RHUMS



Por un lado, que los nodos (términos) que más vinculaciones tienen con el resto tienden a ocupar posiciones centrales en el mapa, mientras que aquellos con menos relaciones se sitúan en la periferia. Este hecho responde a una característica de las representaciones

obtenidas a partir de análisis de redes sociales, en virtud de que desde esta perspectiva de análisis los actores o nodos centrales en la estructura no son necesariamente aquellos que tienen mayor presencia individual (en nuestro caso más alta frecuencia de ocurrencia), sino aquellos que tienen más vinculaciones con los demás (en nuestro caso, co-ocurrencia con más cantidad de términos). En otras palabras, la disposición de los nodos en cada red guarda relación estrecha con la cantidad e intensidad de las asociaciones que cada uno de ellos tiene con el resto.

Por otro lado, los nodos tienden a agruparse en racimos o grupos disciplinares. Los términos que representan a las disciplinas desempeñan un importante rol en la red, actuando como núcleos temáticos alrededor de los cuales orbitan los términos más específicos relacionados más intensamente con ellos. Asimismo, en cada red vemos que hay algunos grupos temáticos que tienen mayor presencia que otros, y esos son los que revelan cual es el perfil temático de la colección. Para facilitar la rápida identificación de los términos que cumplen esa función los hemos resaltado con negrita. En LCNAT (Fig. 1) son: Antropología, Arqueología, Biología, Botánica, Zoología, Geología y Paleontología; en RCNAT (Fig. 2) Antropología, Biología, Geología, y Geología y Geoquímica; en LHUMS (Fig. 3) Literatura, Historia, Filosofía y Lingüística, y en RHUMS (Fig. 4) Literatura, Historia, Filosofía, Educación, Humanidades, Ciencias Sociales, Filología.

Otra característica común a todos los mapas es que el tipo de estructura es más jerárquica que reticular. Salvo contadas excepciones los términos específicos se vinculan bastante poco entre ellos. Ello evidencia sin duda el criterio empleado por las bibliotecas para organizar temáticamente sus colecciones, basado en esquemas jerárquicos del tipo CDU, CDD, u otros semejantes. Que este tipo de esquema se vea reflejado en los mapas no es casual, sino que responde a la metodología que hemos empleado para construirlos, basada en la asociación de un término genérico por cada uno de los términos específicos utilizados para la indización. Asimismo, otro factor asociado a este fenómeno es la cantidad de términos asignados a cada registro. En el caso de los catálogos analizados encontramos que RHUMS tenía un promedio de 2,4 términos por registro; LHUMS una media de 5,3; LCNAT, de 3,5 y RCNAT de 2,5. Es evidente que las revistas son más pobres en su descripción temática, y que cuantos más términos se usen para la indización, mayor será la probabilidad de establecer asociaciones entre ellos. No por casualidad la red de mayor densidad es la de LHUMS.

La dificultad de encontrar relaciones de co-ocurrencia en bases de datos referenciales no es un tema menor para poder aplicar estas técnicas con fines de visualización, y podría explicar el hecho, de que la mayoría de los prototipos de interfaces visuales de sistemas de recuperación de información se hayan planteado para bibliotecas digitales (Marcos, 2004).

En cuanto a las principales diferencias encontradas en los mapas podemos mencionar:

Que aunque es posible identificar que los mapas temáticos pertenecientes a las mismas bibliotecas (LHUMS y RHUMS por un lado, y LCNAT y RCNAT por otro), representan colecciones de temáticas iguales o muy afines, en el caso de CNAT se detecta que los criterios de indización empleados no son los mismos. La diferencia más notoria se puede visualizar en el caso de la Biología, que en RCNAT (Fig. 2) nuclea a la Zoología y a la Botánica, y en LCNAT (Fig. 1) cada uno aparece como un nodo central de un grupo temático bien delimitado.

Para encontrar una explicación a ese fenómeno consultamos a los bibliotecarios encargados de procesos técnicos de la biblioteca a la que pertenecen dichos catálogos. Si bien en ambos casos se utiliza una lista de términos ad-hoc, la diferencia parece estar en que en RCNAT la indización había sido cuidadosamente revisada, normalizando términos, reduciendo la cantidad de entradas diferentes, y respetando siempre un mismo criterio de asignación de descriptores que involucraba a la Biología siempre como término tope de los otros (Zoología y Botánica); y en el caso de LCNAT, la indización se hacía de manera más libre y no se tenía mucho conocimiento sobre cuan coherente y consistente era el criterio utilizado para llevarla a cabo. Por esa razón, en esta última red también aparecen términos sinónimos como Viajes, Expediciones y Expediciones científicas, que bien podrían reunirse bajo una sola forma de entrada, de utilizarse una interfaz visual como el mapa obtenido.

En los catálogos de HUMS parece respetarse un mismo criterio desde la perspectiva de la elección de los términos, y ello posiblemente se deba a que la indización en esta biblioteca se realiza utilizando un tesoro. No obstante, las diferencias radican en la cantidad de términos asignados. Como lo anticipamos en la sección precedente, en RHUMS el promedio de términos por registro es de 2,4 mientras que en LHUMS es de 5,3.

Las distinciones observadas no parecen triviales a la hora de pensar en estos mapas como posibles interfaces de búsqueda. Sin embargo, resulta evidente que el problema está en la indización y no en el método utilizado para la representación visual de la red temática de la colección. Es obvio, que la calidad de los resultados de este tipo de análisis depende de una variedad de factores. Algunos de ellos son la cobertura de la base de datos y la adecuación de los métodos estadísticos escogidos. Sin embargo, el principal factor es la cantidad de términos y la coherencia y la consistencia de la indización (He, 1999), sin lo cual difícilmente pueda obtenerse una buena representación visual de la colección.

Por el contrario, sí son limitaciones inherentes al método, que para obtener mapas legibles y fáciles de interpretar, resulta necesario acotar la cantidad de nodos y enlaces de las redes. Ello significa que no podemos representar la totalidad de términos que describen el fondo documental, sino solo aquellos con mayor presencia o con mayor cantidad de lazos. Ambos criterios son metodológicamente válidos. El criterio de selección de términos por sus propiedades atributivas parte del supuesto de que sólo los nodos con más altas frecuencias de aparición integrarán la red. El segundo circunscribe la población objeto de estudio a los actores que tienen relaciones entre sí (Miceli, 2008). En nuestro caso hemos considerado ambos criterios. Seleccionamos los términos más frecuentes procurando no excluir aquellos con un alto poder de intermediación.

El recorte en la cantidad de términos también hubiera tenido lugar, si en vez de análisis de redes sociales hubiésemos empleado técnicas de *clustering* o escalamiento multidimensional. Como lo hemos comentando previamente, el principal problema radica, en todo caso, en la determinación del punto de corte, que no solo tiene que ver con las frecuencias de ocurrencia sino con el rol que desempeña cada nodo en la red. En este sentido, un claro ejemplo de nodos con una gran capacidad de intermediación son los topónimos que aparecen en el mapa de la Fig. 2. El ejemplo más significativo es el término Argentina, nodo cuya exclusión desvirtuaría considerablemente la estructura temática de la colección de RCNAT. En los otros casos, los nombres geográficos no tenían esa función como tampoco altas frecuencias de ocurrencia, y por tanto quedaron excluidos de los mapas.

Cabe señalar asimismo, que aunque para otros fines los términos con altas frecuencias de ocurrencia suelen ser excluidos por su escaso poder de discriminación (Peat y Willett,

1991), para los objetivos de este estudio dichos términos cumplen una función esencial, en tanto nodos que demarcan los grupos temáticos más representativos de la colección.

7. Conclusiones

Las principales conclusiones a las que arribamos con este estudio son:

Que la combinación de las técnicas de análisis de co-términos y de redes sociales resulta ser una alternativa metodológica válida para la creación de mapas temáticos de las colecciones de las bibliotecas, al menos de bibliotecas universitarias.

Que la metodología resultó ser apropiada tanto para representar colecciones de tamaños, tipos documentales y temáticas diferentes.

Que es posible generar mapas temáticos de las colecciones, independientemente de que las bibliotecas utilicen o no un tesoro para la indización. Asimismo, fue posible comprobar que aún con una cantidad reducida de términos por registro, es posible generar una representación gráfica de la estructura temática de la colección. No obstante, también se evidencia que cuantos más términos se asignen a un documento, mayores serán las posibilidades de establecer asociaciones temáticas entre ellos.

Que aunque no sea posible visualizar en un mapa la totalidad de los términos que describen el fondo documental, el método es viable para reflejar la estructura temática nuclear de la colección, poniendo de relieve los grupos disciplinares más representativos y garantizando legibilidad y facilidad de interpretación.

Por último, creemos que este tipo de mapas, podría servir tanto como una alternativa de interfaz visual para el acceso por materias a las colecciones de las bibliotecas, ofreciendo al usuario una representación simplificada de la cobertura y organización temática de las mismas, como también para detectar problemas en los criterios de indización, y contribuir a la mejora de la calidad de la descripción temática del conjunto documental. Resta ahora poner a prueba si esta estrategia de representación también resulta válida y útil para el usuario final.

Bibliografía

Arntz, R.; Pich, Heribert. Introducción a la terminología. Madrid. Pirámide, 1995

Börner, Katy; Chen, Chaomei, and Boyack, Kevin W. “Visualizing Knowledge Domains”, *Annual Review of Information Science and Technology*, 2003, v. 37, pp. 179-255.

Börner, K.; Chen, C. (2002) “Visual Interfaces to Digital Libraries. Motivation, Utilization, and Socio-technical Challenges”, *Lecture Notes in Computer Science*, 2002, n. 2539, pp. 1-9.

Buchelli Guerrero; Víctor Andrés. “Borrado CERO – Herramienta Informática de vigilancia tecnológica para el análisis de comunidades científicas”. Maestría en Ingeniería de Sistemas y Computación. Universidad Nacional de Colombia, 2006. Consultado en: junio-2008. <http://dis.unal.edu.co/profesores/ypinzon/2013326-206/docs/Tesis0Bucheli.pdf>

Buzydowski, J.; White, H. D. ; Lin, X. “Term co-occurrence analysis as an interface for digital libraries”, *Lecture Notes in Computer Science Series*, 2002, v. 2539, pp. 133-144.

Callon, M. “Éléments pour une sociologie de la traduction, la domestication des coquilles Saint-Jacques et des marins-pêcheurs dans la baie de Saint-Brieuc”. *L'année sociologique*, 1986, v. 36, pp. 169-208.

Caraballo Pérez Y. Los Topic Maps y su relación con las redes sociales. *Acimed*, 2007, v.16, n. 4. Consultado en: junio-2008. http://www.bvs.sld.cu/revistas/aci/vol16_4_07/aci091007.html

Chen, C. (1998) Bridging the gap: The use of Pathfinder networks in visual navigation. *Journal of Visual Languages and Computing*, 1998, v. 9, n. 3, pp. 267-286. Consultado en: agosto- 2008. <http://www.pages.drexel.edu/~cc345/papers/jvlc.pdf>

Drabenstott, Karen; Weller, Marjoric S. “Failure análisis of subject searches in a test of a new design for subject access to online catalogs”. *Journal of the American Society for Information Science*, 1996 v. 47, n. 7, pp. 520-537.

Dürsteler, Juan C. “Visualización de Redes Sociales”. *InfoVis.net*, 2003, n. 136. Consultado en: junio- 2008. <http://www.infovis.net/printMag.php?num=136&lang=1>

Guerrero- Bote, V. P. et al. “Document organization using Kohonen's algorithm”. *Information Processing and Management*, 2002, v. 38, pp. 79-89.

Hassan-Montero, Yusef; Herrero-Solana, Víctor. “Interfaz visual para recuperación de información basada en análisis de metadatos, escalamiento multidimensional y efecto ojo de pez”. *El Profesional de la Información*, 2006, v. 15, n. 4, pp. 278-289. Consultado en: abril-2008. http://eprints.rclis.org/archive/00009547/01/vol15_4.4.pdf

Hassan-Montero, Yusef. “Visualización y recuperación de información”. En: II Encontro de Ciências e Tecnologías da Documentacao e Informacao. Vila do Conde, 27 Abril, 2006. Consultado en: junio- 2008. http://www.nosolousabilidad.com/hassan/visualizacion_y_recuperacion_de_informacion.pdf

He, Qin. "Knowledge Discovery Through Co-Word Analysis". *Library Trends*, 1999, v. 48, n. 1, pp. 133-159.

Herrero-Solana, Víctor; Hassan-Montero, Yusef. "Metodologías para el desarrollo de interfaces visuales de recuperación de información: análisis y comparación". *Information Research: an international electronic journal*, 2006, v. 11, n. 3. Consultado en: abril-2008. <http://eprints.rclis.org/archive/00006616/02/paper258.pdf>

Hildreth, C.R. "The online acces catalog". *Annual Review of Information Science and Technology*, 1985, v. 20, pp. 233-285.

Janecek, Paul; Pu, Pear. "An evaluation of semantic fisheye views for opportunistic search in an annotated image collection". *International Journal of Digital Libraries*, 2005, v. 5, n. 1, pp. 42-56.

Játiva Miralles, Ma. Victoria. "Indicadores de calidad aplicables al análisis, evaluación y comparación de OPACs". *El Profesional de la Información*, 2004, v. 13, n. 1, pp. 28-46.

Kamada, T.; Kawai, S. "An Algorithm for Drawing General Undirected Graphs". *Information Proccesing Letters*, 1989, n. 31, pp. 7-15.

Korn, Flip; Shneiderman, Ben. "MeSHBROWSE: A Tool for Browsing Medical Terms", 1997. Consultado en: setiembre- 2008. <http://hcil.cs.umd.edu/trs/96-01/96-01.html>

Larson, R.R. "Classification clustering, probabilistic information retrieval and the online catalog". *Library quarterly*, 1991, v. 61, n. 2, pp. 133-173.

Latour, Bruno. "Where are the Missing Masses? The Sociology of a Few Mundane Artifacts". En: Bijker, W.E.; Law, J. *Shaping Technology/Building Society: Studies in Sociotechnical Change*. Cambridge: MIT Press, 1992, pp. 151-180. Consultado en: junio-2008. <http://www.bruno-latour.fr/articles/article/50-MISSING-MASSESrepub.pdf>

Lin, X.,; Soergel, D.; Marchionini, G. "A Self-organizing semantic map for information retrieval". En: Proceedings of the 14th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, 1991, pp. 262-269.

Marcos Mora, Mari Carmen. "Algunos principios básicos del diseño de interfaces de usuario" En: Cristófol Rovira; Lluís Codina (dir.). *Documentación digital*. Barcelona: Sección Científica de Ciencias de la Documentación. Departamento de Ciencias Políticas y Sociales. Universidad Pompeu Fabra, 2003. Consultado en: septiembre- 2008. http://www.mcmarcos.com/pdf/2002_algunos-modd.pdf

Marcos Mora, Mari Carmen. "La visualización en el proceso de búsqueda y recuperación de información". En: Rovira, C.; Codina, L. (dir.). *Documentación digital*. Barcelona: Sección Científica de Ciencias de la Documentación del Departamento de Ciencias Políticas y Sociales de la Universidad Pompeu Fabra, 2004. Consultado en: junio-2008. http://www.mcmarcos.com/pdf/2004_visualizacion-modd.pdf

Matthews, J. R. "Public access to online catalogs". New York: Neal-Schuman, 1983.

Mi, Jia; Weng, Cathy. "Revitalizing the Library OPAC: Interface, Searching, and Display Challenges". *Information Technology & Libraries*, 2008, v. 27, n. 1, pp. 5-22.

Miceli, Jorge E. "Los problemas de validez en el análisis de redes sociales: Algunas reflexiones integradoras". *Redes: Revista hispana para el análisis de redes sociales*, 2008, v. 14, n. 1. Consultado en: agosto- 2008. http://revista-redes.rediris.es/html-vol14/vol14_1.htm

Morillo-Velarde Serrano, José. "Aquabrowser en CEUNET: implantación de un sistema de recuperación de la información adaptado a las necesidades de una red de bibliotecas heterogéneas". En: *Proceedings VIII Jornadas de Gestión de la Información*, 2006, pp. 169-180, Madrid (España). Consultado en: setiembre- 2008. http://www.sedic.es/JornadasGI06_Jose_Morillo.pdf

Moya Anegón, F. d.; Herrero Solana, V. "Investigaciones en curso sobre interfaces gráficos en dos y tres dimensiones para el acceso a la información electrónica". *Cuadernos de Documentación Multimedia*, 1999, n. 8. Consultado en septiembre- 2008. <http://www.ucm.es/info/multidoc/multidoc/revista/num8/moya.html>

Moya Anegón, F. d. et al. "A new technique for building maps of large scientific domains based on the cocitation of classes and categories". *Scientometrics*, 2004, v. 61, n. 1, pp. 129-145.

Noyons, E. C. M. ; Moed, H. F.; Luwel, M. "Combining mapping and citation analysis for evaluative bibliometric purposes: a biliometric study". *Journal of the American Society for Information Science*, 1999, v. 50, n. 2, pp. 115-131.

Peat, Helen J. ; Willett, Peter. "The limitations of term co-occurrence data for query expansion in document retrieval systems". *Journal of the American Society for Information Science*, 1991, v. 42, n. 5, pp. 378-383.

Pichinini, M; Martínez, AM.a. "La búsqueda por materia en el catálogo en línea: los problemas del usuario". *Boletín Bibliotecológico de La Plata*, 1996, v. 4, pp. 1-12.

Ríos García, Y. "Catálogos en línea de acceso público: selección bibliográfica". *Revista española de documentación científica*, 1991, v. 14, n. 2, pp. 121-141.

Rodríguez, Joseph A. "Análisis estructural y de redes". Madrid: Centro de Investigaciones Sociológicas, 1995 (*Cuadernos Metodológicos*, 16)

Schneiderman, B. "The eyes have it: A task by data type taxonomy for information visualization". En: *Proceedings of IEEE Symposium on Visual Languages*, 1996, pp. 336-343.

Senso, José A. "Visualizar información: ¿llegó la hora de las bibliotecas?". *El profesional de la información*, 2006, v. 15, n. 4, pp. 248-250. Consultado en: junio-2008. <http://www.elprofesionaldelainformacion.com/contenidos/2006/julio/1.pdf>

Su, S.F. "Dialogue with an OPAC : how visionary was Swanson in 1964?" *Library Quarterly*, 1994, v. 64, n. 2, pp. 130-61.

Vargas Quesada, B. “Visualización y análisis de grandes dominios científicos mediante redes pathfinder (PFNET)”. Granada: Universidad de Granada, 2005. Tesis doctoral.

Vargas Quesada, B.; Moya Anegón, F. d. “Visualizing the Structure of Science”. Berlin: Springer, 2007

Villén Rueda, Luis. “Indización y recuperación por materias en los opacs de las bibliotecas españolas: ¿dos décadas de evaluación?”. *El Profesional de la Información*, 2006, v. 15, n. 2, pp. 87-98.

Wasserman, S.; Faust, K. “Social Network Analysis: Mehtods and Applications”. Cambridge: Cambridge University Press, 1998

White, H. D.; Mccain, K.W. “Visualization of literatures”. *Annual Review of Information Science and Technology*, 1997, v. 32, pp. 99-168.

White, H. D. “Pathfinder Networks and Author Cocitation Analysis: A Remapping of Paradigmatic Information Scientists”. *Journal of the American Society for Information Science and Tehchnology*, 2003, v. 54, n. 5, pp. 423-434.